A Meta-Analysis on the Relationship between Income Inequality and Economic Growth

de Dominicis, L.; Florax, R.J.G.M.; de Groot, H.L.F.

published in
Scottish Journal of Political Economy
2008

DOI (link to publisher)
10.1111/j.1467-9485.2008.00470.x

document version
Publisher's PDF, also known as Version of record

Scottish Journal of Political Economy, Vol. 55, No. 5, November 2008

A META-ANALYSIS ON THE RELATIONSHIP BETWEEN INCOME INEQUALITY AND ECONOMIC GROWTH

Laura de Dominicis, Raymond J. G. M. Florax** and Henri L. F. de Groot****

Abstract

In recent years there is a growing interest in determining the impact of inequality on economic growth. Theoretical papers as well as empirical applications have, however, produced controversial results. Although there is a considerable part of the literature that considers inequality detrimental to growth, more recent studies have challenged this result and found a positive effect of inequality on growth. In this paper, we provide a contribution to the empirical puzzle by using meta-analysis to systematically describe, identify and analyse the variation in outcomes of empirical studies. We find that estimation methods, data quality and sample coverage systematically affect the results. The results point out that it will be particularly useful to increasingly focus research on determining the impact of income inequality on economic growth using single-country data at the regional level, or a relatively homogeneous set of countries with adequate controls for country-wide differences in economic, social and institutional characteristics.

I Introduction

Growing interest in the impact of inequality on economic growth has recently stimulated new theoretical as well as empirical research. Some existing theoretical models propose inequality is detrimental to growth, but alternative theoretical models point at income inequality as an essential determinant furthering economic growth. Benabou (1996) and Aghion *et al.* (1999) provide excellent surveys of the theoretical literature. The line of reasoning in these papers focuses on whether countries will face trade-offs between reducing inequality and improving their growth performance, or instead whether there exists a virtuous circle in which growth leads to lower inequality, and lower inequality in turn leads to faster growth. These divergent theoretical deductions have important policy implications, because stimulating economic growth as

*VU University, Amsterdam, The Netherlands
**Purdue University, West Lafayette, USA, and VU University, Amsterdam, The Netherlands
***VU University, Amsterdam, and Tinbergen Institute, Amsterdam-Rotterdam, The Netherlands

well as obtaining a reasonably egalitarian income distribution is at the heart of the efficiency-equity trade-off that shapes policy discussions in most countries around the world.

The mechanisms linking inequality and growth have also been addressed in an empirical literature (see Campano and Salvatore, 2006, for an excellent review). Early studies are based on the estimation of cross-country growth regressions in which some measure of inequality is added to the set of explanatory variables. Based on this approach, studies such as Alesina and Rodrik (1994), Persson and Tabellini (1994), Clarke (1995), and Deininger and Squire (1998), provide a fairly robust body of evidence for a negative relationship between income inequality and economic growth.[1] Recently, however, the original cross-country evidence is challenged. The availability of data on income distribution for a larger sample of countries and a longer time span has allowed researchers to explore the issue by means of more sophisticated econometric techniques, and frequently evidence is provided for a positive correlation between income inequality and economic growth.

Arguably, the evidence constitutes a theoretical and empirical puzzle; no general consensus has emerged so far. Conclusions seem to depend on theoretical preferences and as far as empirical studies go, on the econometric method employed, the countries considered, and the type of income distribution data used. In this paper, we contribute to the discussion by presenting a meta-analysis of the empirical literature on the relationship between income inequality and economic growth. Traditional approaches of literature review typically use qualitative methods, which provide a chronological, narrative and at times critical overview of the research findings (see e.g. Leoni and Pollan, 2003). In contrast, meta-analysis provides an in-depth quantitative review of the existing literature, and employs formal statistical techniques to summarize the results and to account for differences in study characteristics.

The remainder of the paper is organized as follows. Section II presents a review of the theoretical literature and focuses on the empirical results. Section III provides a description of the sample of studies used in the meta-analysis, and introduces and illustrates the technique of meta-analysis. In Section IV, the potential sources of heterogeneity in the effect sizes are discussed. Section V presents and comments on the results of the meta-regression. Section VI concludes.

II Theoretical Foundations and Empirical Evidence

Theoretical models

The literature on the relationship between inequality and growth dates back to Kuznets (1955). He showed that inequality in per capita income increases

[1] Note that Persson and Tabellini (1994) use an income distribution measure defined in terms of equality among individuals. In their paper the negative relationship between income inequality and economic growth is therefore corroborated by a positive coefficient of the income distribution measure.

in the early stages of development, specifically in the transition from a rural to an industrial organization, and subsequently decreases when the modern structure has penetrated the entire socio-economic texture. The result is the inverted U-shaped relation between inequality and per capita income known as 'Kuznets curve'. This literature was later elaborated upon in several directions. The current literature focuses on the relationship between growth and inequality rather than on the relationship between inequality and the level of economic output, as in Kuznets' seminal work. The latter literature does not provide a unified picture, with some contributions suggesting that inequality is detrimental to growth and others that an initial unequal distribution of resources is a *conditio sine qua non* for the subsequent development of an economy.

The initial literature has brought forward models in which wealth inequality and growth are positively correlated. Aghion *et al.* (1999) summarize the reasons why inequality has been seen to have a stimulating effect on growth in three points. Building on Marxian theories, a first argument is based on the hypothesis that the marginal propensity to save of the rich is higher than that of poor people. If the investment rate is positively related to the savings rate, and investment and growth are positively correlated, more unequal economies are likely to grow faster. Second, in the presence of investment indivisibilities and large sunk costs, the concentration of wealth is essential for the creation of new activities. The third and last argument rests on the trade-off between equity and efficiency, through incentives to workers. If output depends on the work effort of agents, an egalitarian distribution of wages might discourage them from making any additional effort and thus reduce the efficiency of the production system (Mirrlees, 1971).

Following Perotti (1996) theoretical contributions that relate income inequality negatively to economic growth can be grouped into four broad categories: the *endogenous fiscal policy* approach, the *socio-political instability* approach, the *borrowing and investment in education* approach, and the *joint education/fertility* approach.

Within the *endogenous fiscal policy* framework income inequality negatively affects growth through distortions induced by corrective government actions. The more the distribution of income is concentrated within a society, the more the government will tend to introduce distorting redistributive measures, with concurrent contraction of capital investments (often the main target of taxation), and subsequent adverse effects on economic growth. Bertola (1993), Persson and Tabellini (1994), and Alesina and Rodrik (1994) have developed distinct models that establish a trade-off between inequality and growth through the joint effects of a political and an economic mechanism. Redistributive government expenditure and taxation increase as inequality increases, constituting the political mechanism, and as a consequence, growth decreases as fiscal distortions increase, which constitutes the economic mechanism.

According to the *socio-political instability* approach (Alesina and Perotti, 1996) a highly skewed distribution of resources induces people to engage in social activities outside the normal markets, such as crime, revolutions, and violent protests. This, in turn, introduces uncertainty and distrust towards the economic system and discourages investments and capital accumulation. In the long run, it slows down the process of economic growth.

The *borrowing and investment in education* argument stresses the impact of inequality on the ability of individuals to accumulate physical and human capital. Aghion *et al.* (1999) show that if there exist decreasing returns with respect to individual capital investments, and if credit imperfections result in individual investments being an increasing function of initial endowments, inequality is detrimental to growth by concentrating investments among few rich people with a low marginal return to investment. Similarly, Galor and Zeira (1993) prove that in the presence of borrowing constraints the initial distribution of resources matters for the accumulation of human capital. Where the distribution of wealth is highly unequal, and people cannot borrow freely because of capital market imperfections, fewer individuals are able to invest in human capital, and this results in a lower rate of growth.

Education and fertility are two important factors that may affect growth in opposite directions. While investments on education affect growth positively, societies with high fertility rates are often characterized by low rates of economic growth. Within the *joint education/fertility* framework, Galor and Zang (1997) are the first to formalize the link between fertility and schooling decisions and their impact on growth. Given the distribution of income, a higher rate of fertility means that the family has fewer resources to invest in education, with a contracting effect on growth. A theoretical model where the trade-off between inequality and growth works through the channel of fertility decisions demonstrates that economies with a less equitable income distribution experience higher fertility differentials, invest less in human capital, which in turn weakens the process of development (de la Croix and Doepke, 2003).

Recently, Galor and Moav (2004) provided a unified theory in which the relationship between the distribution of income and growth is not stable over time, but depends on the stage of development in a country. The positive impact of inequality upon growth reflects the situation of an economy during its early stage of industrialization. In this phase, the accumulation of physical capital is the principal engine of growth and it is promoted by disparities among individuals. Once the economy has passed over this initial phase, the accumulation of human capital becomes the prime engine of growth and a more equalitarian distribution of resources allows more people to invest in education. In this stage, in the presence of credit constraints, access to education is easier if wealth is evenly spread among individuals, and hence policy decisions have to be directed towards inequality-reducing strategies. Their conclusions are particularly relevant for less developed countries (LDCs). In contrast with the historical growth path of the currently developed countries, where physical capital was the prime engine of growth, human capital accumulation may be the prime engine of growth in some LDCs, even in the early stages of development. In some of the current LDCs, the strong presence of international capital inflows weakens the beneficial role of inequality in stimulating physical capital accumulation. In addition, the tendency to import skill-based technologies in LDCs increases the returns to human capital accumulation and, given credit constraints, strengthens the negative effect of inequality on human capital accumulation, and thus economic growth.

Empirical contributions

In the empirical literature, a similar division between studies reporting a negative effect of inequality on growth and those documenting a positive relationship is apparent. The differences in results can be largely attributed to four factors: differences in data used to measure inequality, the time span of the data, sample coverage and estimation methods.

Income inequality can be measured in several ways. One of the most popular measures is the Gini coefficient.[2] Studies using this inequality indicator largely rely on the data set composed by Deininger and Squire (1996), which contains a much larger and comprehensive sample of data on income distributions than was hitherto available.[3] It contains around 680 high-quality observations (for 108 countries) of the Gini coefficient. To achieve the 'high-quality' standards the data on income have to meet the following requirements: the data must be based on household surveys, the population covered must be representative of the entire country and the measure of income (or expenditure) must be comprehensive, including income from various sources, such as self-employment, non-wage earnings, and non-monetary income. Although the Deininger and Squire data set is not without limitations, it constitutes a substantial improvement in terms of reliability of the data, especially in comparison with the data used in previous studies (see Szekely and Hilgert, 1999; Atkinson and Brandolini, 2001; Knowles, 2005).

Along with the Gini coefficient, researchers have used other measures of inequality such as the share in income of a particular quintile (e.g. the bottom 20 percent), the ratio of incomes of different groups (e.g. the ratio of incomes of the top vs. the bottom of the income distribution), or indicators belonging to the family of generalized entropy measures. Cowell (1995) provides an exhaustive discussion of different inequality measures. However, we prefer to restrict our analysis to studies that use the Gini coefficient to ensure the highest level of comparability among estimates. Moreover, several studies report that virtually all measures of income inequality are highly correlated (Clarke, 1995; Nahuis and de Groot, 2003).

The standard procedure for estimating the impact of inequality on growth is to assume a simple linear relationship, where the logarithmic difference of per capita income at the beginning and the end of the time period is regressed on a

[2] Technically speaking, the Gini coefficient measures the extent to which the distribution of income across households within an economy deviates from a perfectly equal distribution. An index of zero represents perfect equality, while a value of 1 (or 100, depending on scale) implies perfect inequality.

[3] The Deininger and Squire dataset is available via the World Bank at www.worldbank.org. For additional information on the dataset see Deininger and Squire (1996).

number of explanatory variables potentially explaining differences in growth rates of countries, including a measure of income inequality. Specifically,

$$(\ln y_{i,t} - \ln y_{i,t-\tau})\frac{1}{\tau} = \alpha_0 \ln y_{i,t-\tau} + \alpha_1 g_{i,t-\tau} + X_{i,t-\tau}\beta + \varepsilon_{i,t}, \tag{1}$$

where the dependent variable is the average annual growth rate of real GDP per capita, $y_{i,t}$ of country i at time t, τ the time span of the data, g a measure of income inequality such as the Gini coefficient, X a matrix of variables including a constant to allow for other growth-promoting factors, and ε a white noise error term.

Studies based on cross-country regressions typically report a negative and significant relationship between initial income inequality and growth. The negative coefficient holds for different measures of inequality, samples of countries, and time periods. Alesina and Rodrik (1994), Persson and Tabellini (1994), Clarke (1995) and Deininger and Squire (1998) all find support of the existence of a trade-off between the two variables.

One of the main critiques to this kind of regression is that cross-country estimates may be biased due to omitted variables. Factors such as technology, climate, institutions and any other country-specific variable may be important determinants of growth rates and may be correlated with the explanatory variables considered in the model. Many of these factors are typically unobservable. By assuming those factors are constant over time and using longitudinal rather than cross-section data, one can control for unobservable factors using fixed or random effects model. This results in a modified panel data version of equation (1), which reads as:

$$\ln y_{i,t} = \bar{\alpha}_0 \ln y_{i,t-\tau} + \bar{\alpha}_1 g_{i,t-\tau} + X_{i,t-\tau}\bar{\beta} + \xi_t + \upsilon_i + \bar{\varepsilon}_{i,t}, \tag{2}$$

where ξ_t is a time-specific fixed effect, υ_i reflects the characteristics of each country assumed to be constant over time, and $\varepsilon_{i,t}$ collects the remaining part of the error which varies over time and over countries. The choice between various different techniques to estimate (2) is governed by assumptions about the error term and its correlation with the explanatory variables. Standard methods for panel data estimation include the fixed and random effects model. The vast majority of panel data growth studies use the fixed rather than the random effects estimator. The latter requires the country-specific effects υ_i to be distributed independently of the explanatory variables. This requirement is violated by construction for a model like in equation (2), given the dependence of $\ln y_{i,t}$ on υ_i. The fixed effects model allows the unobserved individual effects to be correlated with the conditioning variables. The use of the fixed effects estimator to study inequality and growth does come at a cost. Temple (1999, p. 132) notes that '[. . .] too often researchers use fixed effects approaches to analyse the effect of variables that are fairly constant over time, or that will affect growth only in the long run'. The method also disregards persistent effects and can lead to misleading results in the case where most of the variation is cross-sectional, for instance for a variable such as income inequality. In addition, a problem with both the fixed and random effects estimator is that equation (2) contains a lagged regressor undermining the strict exogeneity assumption of the explanatory variables. In view of these econometric problems many studies resorted to the use of the GMM estimator initially developed by Arellano and Bond (1991). This estimator is based on first-differencing each variable to eliminate the country-specific effects and then uses all possible lagged values of the explanatory variables as instruments to alleviate the problem of endogeneity. However, this approach is also sensitive to critique if used to model the causal relationship between inequality and growth. The use of a lagged explanatory variable in the regression is likely to perform badly when the variables are highly persistent, because in that case lagged levels are weak instruments for first differences (Temple, 1999). The system GMM estimator proposed by Arellano and Bover (1995), and Blundell and Bond (1998) resolves this problem, and has recently been used by Castelló (2004). The idea of the system GMM estimator is to combine a regression in levels with a regression in first-differences in a system of equations. Lagged levels are used as instruments for first-differences, but in addition first-differences are used as instruments for levels.[4]

Some authors claim that the lack of consensus in the empirical findings might be due to the fact that most empirical studies estimate a linear model whereas the correct functional form might be non-linear (Banerjee and Duflo, 2003). They argue that allowing for non-linearity in the effect of inequality, a change in any direction of the income distribution may be detrimental to growth. In addition, doubts have been raised regarding the use of a statistical measure based on income as a proxy for wealth inequality. Alesina and Rodrik (1994), and Deininger and Squire (1998) use data on land inequality in addition to income inequality. Their results, based on a cross-sectional analysis, show that the negative impact of land inequality on growth is more robust than that of income inequality. Castelló (2004) estimates a dynamic panel model where income inequality and human capital inequality are jointly considered. The author finds that the positive correlation between income inequality and growth is robust even when controlling for educational inequality. The relationship between human capital inequality and growth is persistently negative, not only in the long run, as demonstrated in cross-section results, but also in the short term.

A recent study investigates the importance of the shape of the income distribution as a determinant of economic growth and emphasizes how inequality in different parts of the income distribution can affect growth differently (Voitchovsky, 2005). Based on a data set consisting of industrialized countries, the author finds support for the existence of a positive relationship in the upper end of the distribution, while evidence of a negative association is reported for inequality in the bottom part of the distribution.[5]

[4] Monte Carlo experiments (Blundell and Bond, 1998) reveal that this estimator is more robust than the Arellano-Bond estimator in presence of highly persistent series, such as income inequality measures.

[5] The most recent addition to the literature arises out of a special session on income distribution at the 2007 Annual Meeting of the American Economic Association, organised by Dominick Salvatore and documented in a special issue of the *Journal of Policy Modeling* (Salvatore, 2007).

III Meta-analysis

Meta-analysis provides researchers with a useful toolkit to summarize available evidence on a certain topic. Traditional approaches of literature review make use of qualitative methods providing a chronological, narrative and at times critical overview of previously documented empirical research findings. In contrast, meta-analysis provides an in-depth quantitative review of the empirical literature, employing statistical techniques to summarize the empirical evidence.

In 1976, Eugene Glass coined the term *meta-analysis*, and succinctly described it as 'the analysis of analyses' (Glass, 1976). Ever since, a large number of meta-analyses have been carried out in the medical and social sciences, where the experimental setup of the research has made its adoption relatively straightforward. More recently, the meta-analysis technique has proliferated to other fields of research as well. In economics, we find applications in macroeconomics (Stanley, 1998, 2004; Görg and Strobl, 2001; Abreu *et al.*, 2005; Rose, 2005; Dobson *et al.*, 2006), labour economics (Card and Krueger, 1995; Ashenfelter *et al.*, 1999; Longhi *et al.*, 2005), environmental economics (Brouwer *et al.*, 1999) and transport economics (Button, 1995; Button and Kerr, 1996; Button and Rietveld, 2000; Wardman, 2001; Brons *et al.*, 2005), to name just a few. In what follows we apply meta-analytical tools to combine, summarize and analyse the results found in the empirical literature on the effect of income inequality on growth.

In the first stage of the research we conducted a systematic search of the literature via electronic sources. We searched the Economic Literature Index (EconLit)[6] for any reference on 'growth', 'income distribution' and 'inequality'. Furthermore, we reviewed the web pages of institutes engaged in research on income distribution and world poverty reduction.[7] Our search led to about 1800 results. We excluded pure theoretical articles, papers in languages other than English, and we restricted our sample to studies that make use of the Gini index as a measure of income inequality and utilize a linear model linking income inequality to growth. We acknowledge that there are numerous important studies that do not use the Gini coefficient as a measure of income distribution (Persson and Tabellini, 1994; Perotti, 1996), but the restriction we apply guarantees the homogeneity and comparability of the population under investigation.[8] We are also aware that the mixed results found in the literature may be induced by the fact that the majority of the scholars assume a linear relationship between income inequality and growth, while a non-linear

[6] The Economic Literature Index, published by the American Economic Association, provides bibliographic references to a wide range of the economics literature. The number of journals indexed in *EconLit* has grown from 182 periodicals in 1969 to over 400 journals today. In March 2007 the database contained more than 815,000 references.

[7] Specifically, we consulted the World Bank Poverty Net, the Luxemburg Income Studies, and the World Income Inequality Database.

[8] Persson and Tabellini (1994) construct a measure based on the share of the third quintile of the income distribution, while Perotti (1996) uses a combined measure based on the share in income of the third and fourth quintiles in order to better capture what is usually referred to as the middle class.

relationship may be a better specification. In this context, Banerjee and Duflo (2003) presented theoretical and empirical evidence that the relationship between income inequality and growth may be non-linear. To date there are, however, not sufficient studies allowing us to perform a statistical analysis on the evidence of a non-linear relationship between the two variables. Even where a non-linear relationship exists between income inequality and growth '[. . .] it may be argued that linear empirical models are still of considerable value. If interpreted as approximations to an underlying non-linear relationship, conventional growth regressions [. . .] may still provide important information' (Rehme, 2006, p. 28).

After screening the gross list of articles on the above criteria, we were left with 37 studies containing a total of 407 estimates of the coefficient associated with the Gini index. The effect size that we investigate in the meta-analysis is the partial derivative of the average annual growth rate (measured on a per 1% basis) with respect to the Gini coefficient (measured on a 0–1 scale).[9] Table 1 and Figure 1 present some descriptive statistics of the meta-sample. Table 1 shows the composition of our meta-sample in terms of the year of publication, whether the paper is published or not, the number of observations each study contributes to the meta-sample, and the mean value of the estimated effect of income inequality on average annual growth. Slightly more than half of the studies are unpublished working papers, almost all studies (29 out of 37) provide results based on cross-section data and approximately one out of every two studies (15 out of 37, to be precise) give results for panel data (as well), and the average estimated effect size is slightly smaller than zero. Figure 1 illustrates the frequency distribution of the point estimates included in the meta-sample. A remarkable feature is the wide range of the different effect size estimates. Approximately 65% of the estimates are negative (263 out of a total of 407), and approximately 35% (144) are positive.

We now use meta-analytical techniques to further characterize these empirical findings and subsequently identify the heterogeneity across estimates as a function of observable differences in research design and data characteristics, and a random component reflecting unobservable differences across estimates. A first natural question is what the combined estimate of all studies is that adequately represents the true underlying effect size between income inequality and growth. We use two widely used estimators in meta-analysis, which differ in their underlying assumptions.

The *fixed effects* method assumes that there is no heterogeneity among study results and that the different magnitude of the estimates is solely due to sampling variation. Statistically, this is equivalent to the hypothesis that all effect sizes are equal, i.e., $\theta_1 = \theta_2 = \cdots = \theta_k = \theta$, where θ is the true common underlying effect

[9] Suitable transformations are needed depending on the unit of measurement of the growth rate and the Gini coefficient. The reference case we used refers to studies where the growth rate was measured as a dlog of per capita income divided by the length of the period (see equation 1) and the Gini coefficient on a 0–1 scale. Hence, the average annual growth rate is measured on a per one percent basis (e.g. a growth rate of 0.02 identifies the case of an average annual growth rate of 2 per cent). As a result, an effect size of -0.1 implies that an increase in the Gini from 0.3 to 0.4 results in a 1% decrease in the average annual growth rate.

Table 1
Characteristics of the primary studies included in the sample

Study	Outlet: Journal or chapter	Outlet: Working paper	# Obs.	Type data[a]: Cross section	Type data[a]: Pooled	Effect size[b]
Alesina and Rodrik (1994)	+	–	12	+	–	– 0.064
Banerjee and Duflo (2003)	+	–	8	–	+	0.216
Barro (2000)	+	–	6	–	+	– 0.013
Benjamin *et al.* (2006)	–	+	14	+	+	– 0.047
Bleaney and Nishiyama (2004)	+	–	28	+	–	0.013
Castelló (2004)	–	+	13	+	+	0.041
Castelló and Domenech (2002)	+	–	2	+	–	0.036
Clarke (1995)	+	–	7	+	–	– 0.081
de la Croix and Doepke (2003)	+	–	6	+	–	0.017
Deininger and Olinto (1998)	–	+	9	–	+	0.132
Deininger and Squire (1998)	+	–	10	+	–	– 0.027
Figini (1999)	–	+	42	+	–	– 0.058
Forbes (2000)	+	–	10	+	+	0.087
Galor and Zang (1997)	+	–	14	+	–	– 0.051
Gylfason and Zoega (2003a)	+	–	1	+	–	– 0.040
Gylfason and Zoega (2003b)	+	–	1	+	–	– 0.030
Iradian (2005)	–	+	6	–	+	0.180
Keefer and Knack (2002)	+	–	2	+	–	– 0.067
Kenworthy (2004)	+	–	8	+	–	– 0.220
Khoo and Dennis (1999)	–	+	6	+	–	– 0.024
Knell (1999)	+	–	3	+	–	– 0.044
Knowles (2005)	+	–	12	+	–	– 0.020
Larrain and Vergara (1997)	–	+	6	+	–	– 0.098
Li and Zou (1998)	+	–	20	+	+	0.051
Litschig (2005)	–	+	4	–	+	– 0.073
Mbabazi *et al.* (2001)	–	+	15	+	+	– 0.014
Odedokun and Round (2001)	+	–	4	–	+	– 0.084
Panizza (2002)	+	–	34	+	+	– 0.015
Partridge (2005)	+	–	42	+	+	0.096
Persson and Tabellini (1991)	–	+	1	+	–	– 0.055
Rehme (2002a)	–	+	16	+	–	– 0.035
Rehme (2002b)	–	+	19	+	–	– 0.068
Schipper and Hoogeveen (2005)	–	+	2	+	–	0.111
Szekely and Hilgert (1999)	–	+	3	–	+	0.002
Tanninen (1999)	+	–	5	+	–	– 0.138
Voitchovsky (2005)	+	–	14	–	+	– 0.043
Zhu (2001)	–	+	2	+	–	– 0.161
Total	22	15	407	29	15	– 0.016

Notes: [a]Seven studies provide effect size estimates for cross-section as well as pooled data.
[b]Mean effect size per study indicating the effect of a one-unit change in the Gini coefficient (viz. from perfect equality to perfect inequality) on the average annual growth rate (measured on a per one percent basis).

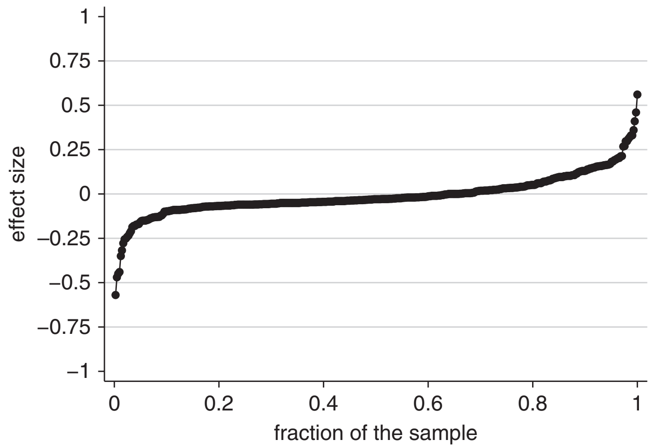

Figure 1. Frequency distribution of the estimated effect of income inequality on growth.

and the subscripts refer to a total of k studies.[10] The pooled fixed effects estimator of the effect size $\bar{T}_{FE}$ is given by:

$$\bar{T}_{FE} = \hat{\theta} = \frac{\sum_{i=1}^{k} w_i T_i}{\sum_{i=1}^{k} w_i}, \tag{3}$$

where i indexes k independent observations, T_i is the estimated effect size, and w_i is the weight assigned to the ith study. Hedges and Olkin (1985) show that the weights minimizing the variance of the statistic are inversely proportional to the square of the standard errors reported in the primary studies (hence the name 'inverse variance' method that is often used to refer to the fixed effects estimator). The weights are given by $w_i = 1/v_i$, where v_i is the estimated variance of T_i. Effectively, the statistic is a weighted average of all effect sizes in the sample, with weights inversely proportional to the precision of the estimates.

The *random effects* method assumes that every study estimates a different effect size, randomly drawn from a larger population with a fixed mean and variance. Under the random effects hypothesis the pooled estimate of the population effect size $\bar{T}_{RE} = \hat{\theta}$ incorporates two components of variation: one is the random variation of the population effect size, and the other is sampling

[10] Note that this statistical theory is originally developed for a single-sampling context often found in experimental studies, where each study provides a single estimate. In our case (and in economics in general), each study oftentimes provides more than one estimate. These estimates are likely correlated because they have been derived utilizing the same or at least a similar dataset. This within-study correlation does not have implications for the bias and consistency of the estimators, but the efficiency is likely overstated. A similar line of reasoning applies if between-study correlation is present, for instance, because of similarities in research design or datasets across studies.

variation. Both are assumed to be normally distributed, with mean zero and variance τ^2 and σ_i^2, respectively. In formal terms:

$$\begin{aligned} T_i &= \theta_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma_i^2), \\ \theta_i &= \theta + \mu_i, \quad \mu_i \sim N(0, \tau^2). \end{aligned} \tag{4}$$

Effectively, the random effects estimator is an inverse-variance weighted average as well, although the weights w_i are now equal to $1/(v_i+\tau^2)$, where v_i represents the within-study variance and τ^2 the between-study variance (see e.g. Sutton *et al.*, 2000, for more details).

We have calculated the pooled fixed and random effect size estimates for the studies in our sample. The fixed effects estimate is equal to zero, albeit not significant, while the random effects estimate equals -0.012, and is statistically significant at the 1% level. Although neither of two models can be said to be 'correct', a substantial difference in the combined effect calculated by the fixed and random effects models may occur when the studies are markedly heterogeneous (Berlin *et al.*, 1989). A Q-test can be performed to check for the adequacy of the null hypothesis of homogeneity of the population effect size. If Q exceeds the critical value of the χ^2 distribution with $k-1$ degrees of freedom, the null hypothesis of homogeneity of the underlying population effect sizes is rejected.[11] Algebraically, the Q-statistic has the following form:

$$Q = \sum_{i=1}^{k} w_i T_i^2 - \frac{\left(\sum_{i=1}^{k} w_i T_i\right)^2}{\sum_{i=1}^{k} w_i}, \tag{5}$$

with all notation as before. In our sample, the Q-statistic equals 3248.2, which clearly rejects the null hypothesis of homogeneity, with a p-value <0.001.

An intensively debated issue in meta-analysis is that of publication bias. The studies for a meta-analysis are usually selected on the basis of a literature search. In these circumstances an inherent selection bias may arise because, for example, studies may tend to be published more readily if they contain statistically significant results, or if they are deemed more 'interesting' in terms of the impact of their outcomes. This bias, often associated with the so-called 'file-drawer effect' because unfavourable results are not published and imagined to be buried in researchers' filing cabinets, is a potentially severe impediment to combining statistical results of studies from the collected literature. Researchers have developed several tools to explore the presence of publication bias, including funnel plots, meta-significance tests, the trim-and-fill method and various other parametric and non-parametric approaches (see Stanley, 2005, for a comprehensive review of the methods available).

[11] As explained in the preceding footnote, strictly speaking the asymptotics of the Q-test are also based on the assumption of independently distributed effect sizes, and the significance level of the test is therefore not fully appropriate in the case of multiple measurements (Sutton *et al.*, 2000).

A popular graphical test for detecting the presence of publication bias is the funnel plot (Egger *et al.*, 1997). The funnel plot compares the effect sizes against some measure of their accuracy, such as sample size or the associated standard error. In theory this plot should depict a 'funnel' shape centred on the true population effect size. Publication bias may lead to asymmetric funnel plots. It is, however, important to realize that publication bias is only one of a number of possible causes of funnel plot asymmetry. The funnel graph in Figure 2 has on the horizontal axis the estimated coefficients associated with the inequality measure and on the vertical axis the associated standard errors. The vertical line in the funnel plot indicates the pooled fixed effects estimate, while the slopes indicate the expected 95% confidence intervals for a given standard error, assuming no heterogeneity between studies. The funnel plot appears not perfectly symmetrical, with a tendency of overrepresentation of results reporting a negative impact of inequality on economic growth.

However, a funnel plot is based upon a subjective and visual inspection of the relationship between the effect size and its precision. Egger *et al.* (1997) proposed a test for detecting asymmetry of the funnel plot. The test detects funnel plot asymmetry by determining whether the intercept deviates significantly from zero in a regression of the standardized effect estimates against their precision. The estimated intercept for our sample is -0.533, with an associate p-value <0.001. Sutton *et al.* (2000) suggest that the resulting slope may be interpreted as a rough estimate of the true effect size after correcting for publication bias. In our sample the coefficient associated with the slope of the regression is close to zero and equal to 0.0003, which is not statistically significant (p-value = .53).

From the above results we infer the following preliminary conclusions. First, by pooling the estimates in our sample and assuming that all variation across effect sizes is purely random and unobservable as in the random effects model,

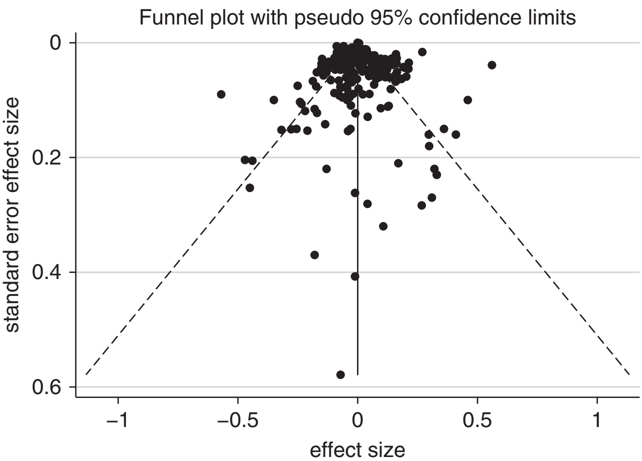

Figure 2. Funnel plot of 407 estimates of the impact of income inequality on growth.

we have to accept the hypothesis that income inequality and growth are negatively correlated. However, after testing for publication bias, we observe that negative values are overrepresented in the collected literature, and concluding in favour of a trade-off between inequality and growth may therefore be misleading. In addition, the results of the Q-test suggest that there is heterogeneity in terms of the underlying population effect sizes, but so far we have treated this variation as unobservable and random. However, part of this variation may be structurally associated with differences across studies and estimates that we can very well observe as identifiable characteristics of the primary studies, including their data characteristics, geographical coverage, estimation procedures, and study design. In the next section, we therefore turn to a multivariate analysis to investigate whether the variation found in the literature can be adequately represented using a combination of observable differences across studies and random variation in the population effect sizes. We will then also allow for within-study correlation of estimates derived from the same study.

IV Meta-regression

Traditional estimators in meta-analysis

Meta-regression is an adequate tool to model the heterogeneity in findings of a body of studies. Technically speaking, it is a regression where 'the dependent variable is a summary statistic, perhaps a regression parameter, drawn from each study, while the independent variables may include characteristics of the method, design and data used in these studies' (Stanley, 2001, pp. 132–3).

Two regression models widely used to control for heterogeneity in study results are the *fixed effects* model and the *mixed effects* model. The fixed effects model assumes that the variability among the effect sizes can be fully explained by a series of s moderator variables that account for differences in study characteristics:

$$T_i = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_s x_{s,i} + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma_i^2). \tag{6}$$

In comparison with the pooled fixed effects estimator that we used in the preceding section, the null hypothesis of one single effect size is relaxed through the assumption that the effect sizes can vary according to predefined differences in study characteristics, but we still assume that all the variation is systematic and fully predictable by a number of covariates.[12]

[12] The estimation is performed by means of weighted least square algorithms, with the weights inversely proportional to the precision of the estimates (i.e., the square of the standard errors reported in the primary studies). It is important to note that weighted least squares performed in standard statistical packages is based on a model that is slightly different from the meta-analysis fixed effects model. While the coefficients are still valid, the reported standard errors require to be adjusted according to:

$$SE_{adj} = SE_{WLS}/\sqrt{MSE}$$

where SE_{adj} is the adjusted standard error, SE_{WLS} the standard error as reported by the computer program, and MSE is the mean squared error from the analysis of variance for the

In the mixed effect model, the variability beyond observation-specific sampling error is derived partly from systematic factors (as in the fixed effects model), and partly from random sources μ_i which are assumed to be distributed with mean zero and variance τ^2:

$$\begin{aligned} T_i =&\beta_0 + \beta_1 x_{1,i} + \cdots + \beta_s x_{s,i} + \varepsilon_i + \mu_i, \\ &\varepsilon_i \sim N(0, \sigma_i^2), \quad \mu_i \sim N(0, \tau^2). \end{aligned} \tag{7}$$

The mixed effects model allows for the presence of heterogeneity by assuming that the underlying effects follow a normal distribution around the mean effects predicted by the covariates. Equation (7) comprises two error components, ε_i and μ_i, which are jointly considered by additively incorporating the variances of the random terms in the weights. Effectively, the mixed effects estimator allows for within- and between-study variance. Estimation is based on an iterative restricted maximum likelihood estimator (see Sutton *et al.*, 2000, for details).

It is often times debatable whether a fixed or a mixed effect model is the most appropriate to investigate the heterogeneous distribution of a given sample of effect sizes. The fixed effects model is quite restrictive, as it assumes that all heterogeneity is perfectly observable. In spite of that, the fixed effects model has more statistical power to identify systematic between-study differences. The mixed effect model relaxes the assumption that all heterogeneity is fully observable, but at the cost of statistical power in identifying moderator effects. The mixed effect model has a further drawback, in that it assumes additivity of the effect size's variances, which is related to the requirement that the effect sizes are independently distributed.

The hierarchical linear regression model

The data used in the meta-analysis are characterized by an inherently hierarchical structure, with observations clustered within studies. The models presented above all assume that the estimated effect sizes are independently distributed, regardless whether they are taken from the same study or not. If primary studies report multiple measurements of the effect size, the assumption that the observations are independent replications of a stochastic process can be easily criticized. The existence of a non-zero within-study correlation implies that conventional estimation procedures (such as OLS) lead to incorrect inferences (Goldstein, 1995). In particular, the estimated parameters are not correct and especially when the within-study correlation is significant the conventional regression procedure will tend to underestimate the standard errors of the coefficients (Bateman and Jones, 2003).

Researchers have generally adopted two approaches to deal with multiple measurements sampled from the same study. The first approach considers the use of a single effect size value for every study. The value can be an aggregate

regression (see e.g. Sutton *et al.*, 2000, p. 94). The adjustment of the standard errors is needed because of the assumption that the variation across effect size estimates is fully attributable to observed differences in the characteristics of studies.

statistic (i.e., average, median) or, alternatively, one observation randomly selected from the entire set of estimates of a particular study. This method can be criticized for not using all the information contained in the primary studies. In addition, the selection criterion can only be determined on subjective grounds. In the second approach, all measurements are individually included in the analysis and treated as weighted independent replications, with weights proportional to the number of estimates contained in the study (Rosenthal, 1991, p. 27). The weighting scheme makes it possible to account for the fact that studies with many measurements may have a larger impact on the results. However, neither of these approaches explicitly considers the hierarchical nature of the data. Multilevel linear models (Goldstein, 1995) can be applied to control for the presence of within-study dependence. This class of models has been frequently used in epidemiology and education research in which the clustering of the units within groups is usually evident (e.g. patients are nested within treatment groups, students are nested within schools).

Raudenbush and Bryk (2002) have pointed out that meta-analysis may be viewed as a special case of the two-level hierarchical model. In each study, a within study model is estimated, and a second level, or between study model, is added to explain the variation in the within study parameters as a function of differences between the studies. So far, multilevel models have been used in meta-analysis mostly in the field of health studies (Beacon *et al.*, 1999; Rutter and Gatsonis, 2001) and education (Goldstein *et al.*, 2000). In economics very few meta-analysis have been carried out in which the hierarchical structure of the data is explicitly incorporated in the regression model (Bijmolt and Pieters, 2001; Brouwer *et al.*, 1999; Bateman and Jones, 2003).

In a simple two-level hierarchical regression framework, a general model for meta-analysis with multiple measurements within studies leads to the following formulation, where the effect size is regressed on a set of explanatory variables plus an error term that now consists of two distinct components:

$$\begin{aligned} T_{ij} &= \beta_{0j} + \beta_1 x_{ij}^1 + \cdots + \beta_K x_{ij}^K + \varepsilon_{ij}, \\ \beta_{0j} &= \beta_0 + \mu_{0j}, \end{aligned} \tag{8}$$

where $\varepsilon_{ij} \sim N(0, \sigma^2)$ and $\mu_{0j} \sim N(0, \tau^2)$, the i's are the individual observations nested in study j, β_0 is a constant, x_{ij}^k are K explanatory variables ($k = 1, 2, \ldots, K$), ε_{ij} represents the error term at measurement level, and μ_{0j} is the error term at the study level. This two-level hierarchical model is a mixed random effects model, which allows the constant to randomly vary across studies.[13] Note that the specifications in equations (7) and (8) are rather similar. Both the mixed effects regression model and the two-level hierarchical model assume that the moderator variables chosen by the researcher are not fully able to account for variation in the estimated effect sizes. However, whereas in the mixed-effects regression model observations in the sample are assumed to be independent, the

[13] See Goldstein (1995) for a methodological discussion of the two-level hierarchical model. In particular, the inadequacy of the OLS estimator in presence of intra-unit correlation is formally discussed in Section 2.8.

same is not true for the hierarchical model, where it is assumed that observations within the same study are dependent, leading to a nested structure of the error term as in equation (8).[14]

In the next section, we present the results of the meta-regression. The estimates reported in the primary studies are regressed over a set of moderator variables chosen in such a way that they control for observable differences in study characteristics. The potential dependence problem induced by the presence of multiple measurements of the effect size from the same study is explicitly taken into account by imposing a nested structure of the error term as in equation (8). In order to give more importance to more accurate estimates, we weight the units at level 1 in the hierarchical level model with weights defined as the reciprocal of the sampling variance. For the sake of comparability, the results of the mixed effect model are also reported.

V Meta-regression Results

Table 2 presents the results of the regressions using the mixed effect model and the hierarchical linear model. We use the latter results for the discussion below. Note that the estimated coefficients are similar in the two models. The estimated standard errors are different due to the within-study correlation. For ease of interpretation, the dependent variable is now defined as the partial derivative of the growth rate measured as a percentage (rather than on a per one percent basis) with respect to the Gini-coefficient, measured on a 0–1 scale.[15]

In a multi-level model, the model fit is assessed using a likelihood ratio test, based on the log-likelihood difference between the full nested model (the hierarchical model where no explanatory variables are added) and the nested model containing only a constant (Goldstein, 1995). The likelihood ratio equals 416.10 and is highly significant. Hence, in presence of intra-unit correlation, the full nested model provides a better fit. Below, we discuss to what extent the heterogeneity in the empirical results can be attributed to differences in the type of data used and the estimation method, data quality, the time period considered, and sample coverage.

Studies based on cross-country regressions typically report a negative and significant relationship between initial income inequality and growth. The finding of a negative coefficient is robust against different measures of inequality, and different samples of countries and time periods. Alesina and Rodrik (1994), Clarke (1995) and Persson and Tabellini (1994) all find support of the existence of a trade-off between the two variables. Using panel data models, the existence of a negative influence of inequality on growth is refuted. Li and Zou (1998), Forbes (2000) and Deininger and Olinto (1998) all use panel data over 5-year intervals and find evidence of a positive and significant impact

[14] Bateman and Jones (2003) note that in terms of model interpretation, it is the stratification of the error term to form the random parameters in a two-level hierarchical model that differentiates a multilevel model from more traditional regression analysis techniques.

[15] As compared to Figure 1, the effect size in the regression analysis is thus a factor 100 larger.

Table 2
Meta-regression results[a]

Moderator variables	Mixed effects model	Hierarchical linear model
Constant	– 50.890***	– 53.649***
	(5.153)	(15.694)
Structure of the data (cross-section)		
Pooled	1.226**	0.932
	(0.550)	(2.009)
Estimation method (OLS)		
Fixed effects	0.086***	0.079***
	(0.031)	(0.012)
Random effects	0.022	0.016
	(0.040)	(0.010)
Endogeneity	– 0.885**	– 0.978
	(0.412)	(1.765)
Characteristics of data on income distribution		
Quality of data (high quality)		
High and low	– 1.645***	– 1.212
	(0.290)	(1.069)
Low	– 5.531***	– 5.144***
	(0.434)	(1.884)
Dataset on income distribution (Deininger and Squire)		
Other data set	0.228	0.453
	(0.383)	(1.092)
Other data set × cross section	2.529***	1.660
	(0.382)	(1.139)
Dynamic of the inequality index (Initial value)		
Average value	2.712***	2.274
	(0.377)	(2.026)
Definition of the Inequality measure, based on (Income)		
Expenditure	– 3.016***	– 2.633***
	(1.012)	(0.789)
Adjusted	– 0.656**	– 0.573
	(0.294)	(0.744)
Mixed	– 0.012	– 0.008
	(0.031)	(0.012)
Sample of countries (OECD countries included)		
Exclusively LDCs	– 1.780***	– 1.710**
	(0.294)	(0.856)
Geographical aggregation level (countries)		
Regions	– 3.269***	– 3.872***
	(0.442)	(1.393)
Time horizon		
Length of the growth period†	– 0.273***	– 0.257***
	(0.024)	(0.093)
Initial year†	– 0.051***	0.025
	(0.009)	(0.040)
Conditioning variables		
Other definition of inequality included	2.177***	1.756**
	(0.344)	(0.860)
Regional dummies included	2.773***	2.046*
	(0.244)	(1.191)

Table 2 (*Continued*)

Moderator variables	Mixed effects model	Hierarchical linear model
Publication characteristics		
Year of publication†	0.523***	0.566***
	(0.054)	(0.165)
Working paper (journal or book)	– 5.019***	– 4.473***
	(0.336)	(1.347)

Notes: [a]The dependent variable is defined as the partial derivative of the growth rate (measured as a percentage) with respect to the Gini-coefficient (measured on a 0–1 scale). The omitted category for dummy variables is provided in brackets, and standard errors are reported in parentheses. The statistical significance of the parameters is indicated by ***, **, and *, referring to the 1, 5, and 10% level, respectively. All moderator variables enter the regression equation as dummy variables, except those labelled with a †, which are continuous variables. The variable 'Year of publication' equals the year of publication minus 1900. The number of observations in the mixed effects model is 407. In the hierarchical linear model, the number of level-1 observations equals 407, and the number of level-2 observations is 37.

of inequality on growth.[16] However, the statistically significant coefficient disappears when the system GMM estimator is adopted (Castelló, 2004). In addition, Barro (2000) has heavily criticized the fixed-effect estimator, because the estimator exacerbates the bias due to the measurement error.[17]

The meta-regression results show that the coefficient for the dummy variable labelled *pooled* (with one for pooled data, and zero for cross-section data) is positive, albeit not significant. Contrary to the bivariate results reported above, the results in a multivariate setting with appropriate controls for study characteristics show that studies utilizing pooled or cross-section data do not tend to lead to different conclusions about the relationship between income inequality and growth. The differences in the estimates may however be due to other factors, for instance the estimation technique, rather than to the structure of the data themselves.

We therefore also control for the use of different estimation methods in the primary studies. We use estimates based on OLS as reference category to evaluate the impact of alternative estimators on the effect sizes. *Fixed effects* and *Random effects* are dummy variables equal to one when the primary study uses a fixed or random effects panel data estimator, respectively. The results show that only the effect sizes estimated with a fixed effects estimator are significantly different from OLS results. They are on average 0.079 percentage points higher. The result is highly significant even after controlling for the within-study

[16] While Li and Zou (1998) use a fixed effects model to estimate the relationship, Forbes (2000) and Deininger and Olinto (1998) estimate a panel data model using the GMM estimator proposed by Arellano and Bond (1991).

[17] Using panel data over 10-year intervals and a three-stage least square estimator which treats country specific effects as random, the author finds no significant relationship between inequality and growth in the whole sample. However, after considering separately the relationship for rich and poor countries, he finds evidence of a negative relationship for poor countries and a positive relationship for rich countries. Note, however, that Barro (2000) obtains a positive coefficient for rich countries only after including an additional explanatory variable for fertility. When the fertility variable is omitted, the point estimate is negative although statistically insignificant.

dependence by means of the hierarchical model. The positive effect of inequality on growth is thus unlikely to be attributable to the use of pooled data as such, but it is more likely associated with the use of a fixed effects estimator, which is not undisputed because inequality tends to be highly persistent over time (see Partridge, 2005).

A potentially serious problem in growth regression is endogeneity. It is plausible to assume that income inequality is jointly determined with the rate of economic growth. If the independent variables are endogenous and thus correlated with the error term, the OLS/fixed effects estimators are biased and inconsistent. The fact that the explanatory variables are dated at the beginning of the growth period reduces the problem of endogeneity. However, when the variables are highly persistent – as in the case of the Gini coefficient – the problem of endogeneity may still persist. *Endogeneity* is a dummy variable equal to one when the primary study controls for the potential problem of reversal causality by using instrumental variables in cross-section studies or GMM estimators in studies based on pooled data. The meta-regression results show that the effect of using IV or GMM estimators is negative, but it is not significantly different from zero.

Another common concern in the literature is that data on income distribution are likely to be undermined by measurement error. When a variable is badly measured its coefficient is biased towards zero (i.e., the so-called *attenuation* effect) resulting in a weaker impact on the dependent variable. In multivariate regression models the consequences are even more serious. The error in one of the dependent variables not only affects the coefficient of the variable itself, but the other coefficients are biased as well, although in an unknown direction (Greene, 2000, p. 86). We addressed the problem associated with the reliability of the data by looking at the quality of the data on the income distribution used in the primary studies. We use two different dummies to indicate the use of data of *low* quality, and data of *high and low* quality. For the interpretation of the results, we used studies that use measures of income distribution obtained exclusively from data of high quality as the reference category. The results show that when the study uses data of mixed quality, the estimated effect sizes are not significantly different from the ones obtained when the study uses only data of high quality. Moreover, when the quality of the data is low, the estimates are significantly different from those obtained when only high quality income data are used in the primary studies. The coefficient associated with the use of poor-quality data is negative, and relatively big in magnitude (-5.144), and significant at the 1% level. If it is assumed that data of low quality generally refer to LDCs, or to studies in which LDCs are strongly represented, we find confirmation to the hypothesis presented in Galor and Moav (2004) that the rate of growth is more likely to depend negatively on income inequality in less developed economies, because of the important role played by human capital accumulation as prime engine of growth in these economies, also in the early stages of their economic development.

We also analysed the effect of the use of different data sets on the effect on income inequality on economic growth. The reference category is the use of the data set compiled by Deininger and Squire, DS for short. As we would like to check whether the adoption of a different data set provokes a bias towards zero in the estimates, we have analysed the issue separately for the estimates obtained through cross-section data, and estimates based on pooled data. *Other data set* is a dummy equal to one when the primary study does not use the data set compiled by DS. The estimated coefficient is not significantly different from zero. This result is not surprising if we consider that the majority of the studies in our sample which do not use the DS data set, obtain information on the income distribution from sources that are similar in terms of quality to the information contained in the database developed by DS (i.e., the *LIS – Luxembourg Income Study* database or the *WIDER – World Income Inequality* database).[18] We also define a dummy variable labelled *Other data set* $\times$ *cross-section* which is equal to one when the primary study uses cross-section data and derives information on the income distribution from a data set different from the DS data set, and zero otherwise. This variable allows us to check whether there are significant differences in the estimated inequality effect for studies using data sets before the DS data set becoming available. Note that all studies performed before the publication of the DS data set make use of cross-section data. The estimated coefficients for *Other data set* and *Other data set* $\times$ *cs* show that there is generally no heterogeneity in the estimated effect of income inequality on economic growth due to the adoption of a database other than the data set compiled by DS.

The paucity of long time series on income distribution has induced researchers to replace the measure of inequality measured at the beginning of the period with the average value over the entire period. Experience suggests that within a country inequality does not change substantially over time. The variability of the index is even lower in the case of pooled data, where period averages are rarely longer than 5 years. Our hypothesis is that the use of the initial or average value should not affect the outcomes. The variable *average value* is a dummy variable that is equal to one when the primary study uses period averages of the Gini index. The reference category is the adoption of the index at the beginning of the period. Although positive, the estimated coefficient is not significantly different from zero.

Atkinson and Brandolini (2001) and Knowles (2005), among others, have cast doubt on the cross-country comparability of data on income distribution for most of the existing data sets. Inequality can be measured in several ways, using data based on pre-tax income, post-tax income, and expenditure. Even the units of measurement may differ. Some surveys use individual data, while other surveys consider households data. Owing to the lack of comparable data for a sufficiently big group of countries, researchers have been forced to use data on income distribution with different specifications. Some authors suggest transforming the original data to achieve a higher level of cross-country

[18] The WIDER database is a secondary datasource containing data from the Deininger and Squire database, the Luxembourg Income Study, and other new sources as they become available.

comparability (Deininger and Squire, 1996; Perotti, 1996; Barro, 2000). A potentially serious problem in the empirical literature examining the impact of income inequality on growth is that, with only few exceptions, studies tend to use measures of income distribution that are not measured in a consistent manner. For instance, data on expenditure are likely to produce lower levels of inequality than income data. Data on gross income and net income may lead to different conclusions on the impact of inequality on growth.[19] We therefore check whether the use of different specifications of income affect the magnitude of the estimated effect size. The reference category refers to studies where the inequality measure is based on income data.[20] We define three control variables, referring to studies where the measure of inequality is obtained using: (i) expenditure data, (ii) adjusted income inequality measures (usually the adjustment method suggested by DS), and (iii) a mixed measure (that is, studies in which the author specifies that the measures of income distribution are based on different income definitions). Table 2 shows that only the coefficient on the dummy variable *Expenditure* is significantly different from zero, showing that inequality measures based on expenditure data are associated with smaller estimated effects of income inequality on economic growth. The dummy *Adjusted* is used here to control for the fact that the study transforms the data to increase the level of comparability. In our database, this type of correction is mostly performed using the adjustment procedure proposed by DS.[21] The resulting coefficient is negative, but not significantly different from zero.

We also investigate the consequence on the parameter estimate of using different samples of countries. The reference category is given by studies that include only OECD countries or both OECD and LDCs in their sample. Barro (2000) and Galor and Moav (2004) argue that the relationship between inequality and growth is different in rich and poor economies. The variable *LDCs* is a dummy variable that takes on a value of one when the primary study includes only LDCs. The coefficient is negative and significant at the 5% level, lending support to the hypothesis that especially for less developed economies income inequality hampers subsequent growth.

Another interesting issue to investigate is whether the geographical level at which the analysis is conducted influences the results. We already discussed the problem associated with studies where data on income are not defined in a homogenous way. This problem may be partly solved by looking at the impact of inequality on growth within a country. Such an analysis can be based on differences between individuals or territorial units in a country. While there are significant differences across countries in the way data on income are collected,

[19] Rehme (2002a) notes that post-tax income depends both on pre-tax income and on redistribution. As a result, when a study mixes gross and net income, the coefficient associated with the inequality measure may be picking up not only the impact of income distribution on growth, but also the impact of redistribution on growth.

[20] We cannot distinguish between net or gross income, because this is not always clear from the information provided in the primary studies.

[21] They suggest adding 6.6 points when the indexes are not based on income, but rather on expenditure data.

these differences are likely to be less important in surveys conducted in a particular country across time (Iradian, 2005). The analysis of the growth-inequality linkage on a regional basis may therefore be more informative than the analysis based on worldwide cross-country data. Moreover, some of the variation observed in the cross section of countries tends to vanish after allowing for regional effects. We define a dummy variable labelled *Regions* to record whether studies estimate the partial correlation between inequality and growth across countries, or across regions in the same country. We find evidence that, on average, studies at regional level report estimated coefficients that are significantly lower than those obtained through cross-country regressions.

Another hypothesis concerns the impact of the length of the growth period on the magnitude of the estimated effect sizes. If it is true that the negative effect of inequality on growth holds only in the long run, an increase in the length of the growth episode is likely to produce estimates that are on average lower. The variable *Length of the growth period* is a continuous variable that includes the number of years over which the growth rate is calculated. What we find is the longer the growth period, the lower the coefficient associated with the Gini index. This result is not fully corroborated by the results obtained when we control for the initial year in the growth regression. The coefficient associated with the moderator variable *Initial year* is positive – indicating that on average the higher the initial year, and thus the shorter the growth period, the higher are the estimated effect sizes of the primary studies – albeit not significant.

Furthermore, we checked for the impact of two types of conditioning variables widely used in this literature. First, we controlled for the adequacy of a measure based on income to describe the relationship between inequality and growth. Deininger and Squire (1998) show that while the index based on income is not significant after the inclusion of regional dummies, the same is not true when inequality is measured in terms of asset distribution, for which the negative coefficient holds and is significant even after controlling for regional effects. Castelló (2004) estimates a regression where a measure of human capital inequality and a measure of income inequality are considered simultaneously in the regression. The author finds that while income inequality is not robust to the econometric specification adopted, the negative relationship between human capital inequality and growth persists also when panel data estimations are carried out. We explore the effect of the introduction of the inclusion of additional measures of inequality in the primary regression. *Other definition of inequality* is a dummy variable equal to one when the primary study uses more than one definition of inequality in the regression, and zero otherwise. If it is true that income inequality is not fully capturing the impact of an unequal distribution of resources on growth, we expect the impact of income inequality on economic growth to be weaker when the study includes another explanatory variable that controls for the effect of other determinants of inequality. The meta-regression results show that the associated coefficient is positive and significant. If we consider that in our database the average value of the effect size is negative, a positive value of the dummy variable *Other definition of inequality* indicates that, *ceteris paribus*, when the study includes a measure of inequality that is not income, the coefficient associated with income inequality becomes smaller.

Another hypothesis concerns the inclusion of regional dummies in the primary studies. The use of country dummies is in some part capturing the effect of country-specific attributes (history, factor endowments, and technological differences) in a way similar to the fixed effects model. *Regional dummies* is a variable that takes on the value of one when the primary studies incorporate regional dummies in the base-regression. The meta-regression results show that the coefficient associated with using regional dummy variables is positive and significant. Similarly to the use of a fixed-effects estimator in the primary studies, the inclusion of country-specific dummies produces estimated effect sizes that are bigger in absolute value. One should note that the effect of including regional dummies is substantially bigger as compared with the effect of using a fixed effects estimator.

Finally, we checked whether systematic reporting trends are present in this literature, either according to the year of publication, or to the type of publication. The variable *Year of publication* is used to control for the presence of a time trend in the publication of the results. The meta-regression results show that resulting coefficient is positive and significantly different from zero. Differences in study results may also be associated with differences in publication outlets, specifically working papers vs. journal articles and book chapters. We therefore define a dummy variable labelled *Working paper*, which takes on a value equal to one if the estimate comes from an unpublished manuscript, and a value of zero if the estimate is from a paper published in a journal or in a book. The resulting coefficient is negative, relatively big in terms of magnitude, and highly significant. This result should, however, be interpreted cautiously, because the majority of studies in our database that are unpublished estimates the relationship between inequality and growth using cross-section data. We may therefore partly be picking up the effect linked to the structure of the data.

VI Conclusions

Economic theory does not unambiguously predict the direction of the effect of income inequality on economic growth. However, the direction and magnitude of the association between income inequality and economic growth are important for policy decisions and policy evaluation. This explains the vast number of empirical studies devoted to this issue. In the empirical literature, the majority of cross-sectional studies has found a negative relationship between income inequality and growth. However, the negative effect seems to disappear when the models are estimated using panel data techniques. So far no clear conclusion has been reached, giving support to the critique that income inequality might be a poor proxy for wealth inequality, as well as casting doubt on the quality of the data used in the analyses. The present paper addresses these issues, and investigates their impact. Specifically, we provide a quantitative analysis of the empirical literature on income inequality and growth using the tools offered by meta-analysis. In this conclusion, we summarize our main findings, which may be of help for future research in this field.

The results of the meta-analysis show that it is misleading to simply speak of a positive or negative relationship between income inequality and economic growth. Differences in estimation methods, data quality and sample coverage substantially affect the magnitude of the estimated effect of income inequality on economic growth. In particular, studies that use a fixed effects model systematically report higher estimates of the effect size. The use of fixed effects estimators as well as the use of regional-specific dummies reduce the negative impact of inequality on growth in cross-section estimates, and accentuate the positive effect in studies based on pooled data. We also find that the negative impact of an uneven distribution of income on economic performance is larger in LDCs. In addition, the length of the growth period that is considered has an important influence on the outcomes. The longer the length of the growth period, the lower the magnitude of the coefficients found in the studies. This result gives support to the hypothesis that the mechanism at the basis of the relationship between inequality and economic growth works differently in the short and in the long run. We also find that the quality of the data on income distribution is a crucial factor. When authors use data of relatively poor quality, the impact of inequality on growth (either positive or negative) is weaker. In addition, the magnitude of the estimates is significantly affected by the inclusion of additional measures of inequality (i.e., land inequality or human capital inequality). An important result of our study relates to the use of consistent data on inequality. We find that when the inequality measure is based on expenditure data the estimates in the primary studies tend to differ from those obtained for income-based data. Apparently, the use of inequality measures based on a mix of income definitions (pre-tax, post-tax, expenditure) does not produce estimates that are significantly different from those using only income data.[22]

These conclusions provide clear pointers to decisions researchers have to make when considering the research design of a new study on income inequality and economic growth. In this respect, it is particularly promising if attention would shift towards samples of regions within one country, or a limited set of countries with similar characteristics, or alternatively with different characteristics to the extent that these can be controlled for in the specification of the regression model.

Acknowledgements

The authors would like to thank the editor and two anonymous referees for useful comments and suggestions. Any remaining errors are ours. The first author acknowledges financial support from the European Union's Sixth Framework Programme for Research and Technological Development.

[22] For a different conclusion and a valid discussion on this aspect see Rehme (2002a).

References[23]

Abreu, M. A., de Groot, H. L. F. and Florax, R. J. G. M. (2005). A Meta-analysis of β-convergence: the legendary 2%. *Journal of Economic Surveys*, **19**, 3, pp. 389–420.

Aghion, P., Caroli, E. and Garcia-Penalosa, C. (1999). Inequality and economic growth: the perspective of the new growth theories. *Journal of Economic Literature*, **37**, 4, pp. 1615–60.

Alesina, A. and Perotti, R. (1996). Income distribution, political instability, and investment. *European Economic Review*, **81**, 5, pp. 1170–89.

Alesina, A. and Rodrik, D. (1994). Distributive politics and economic growth. *Quarterly Journal of Economics*, **109**, 2, pp. 465–90. ■

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, **58**, 2, pp. 277–97.

Arellano, M. and Bover, O. (1995). Another look at the instrumental-variable estimation of error-components models. *Journal of Econometrics*, **68**, pp. 29–51.

Ashenfelter, O., Harmon, C. and Oosterbeek, H. (1999). A review of estimates of the schooling/earnings relationship, with tests for publication bias. *Labour Economics*, **6**, pp. 453–70.

Atkinson, A. and Brandolini, A. (2001). Promise and pitfalls in the use of 'secondary' datasets: income inequality in OECD countries as a case study. *Journal of Economic Literature*, **39**, 3, pp. 771–99.

Banerjee, A. V. and Duflo, E. (2003). Inequality and growth: what can the data say? *Journal of Economic Growth*, **8**, pp. 267–99. ■

Barro, R. J. (2000). Inequality and growth in a panel of countries. *Journal of Economic Growth*, **5**, pp. 5–32. ■

Bateman, I. J. and Jones, A. P. (2003). Contrasting conventional with multi-level modeling approaches to meta-analysis: expectation consistency in U.K. woodland recreation values. *Land Economics*, **79**, 2, pp. 235–58.

Beacon, H., Clarke, R. and Frost, H. (1999). Use of hierarchical models for meta-analysis: experience in the metabolic ward studies of diet and blood cholesterol. *Statistics in Medicine*, **18**, pp. 1657–76.

Benabou, R. (1996). Inequality and growth. NBER Working Paper no. 5658.

Benjamin, D., Brandt, L. and Giles, J. (2006). Inequality and growth in rural China: does higher inequality impede growth? IZA Discussion Paper Series no. 2344. ■

Berlin, J. A., Laird, N. M., Sacks, H. S. and Chalmers, T. C. A. (1989). Comparison of statistical methods for combining event rates from clinical trials. *Statistics in Medicine*, **8**, pp. 141–51.

Bertola, G. (1993). Factor shares and savings in endogenous growth. *American Economic Review*, **83**, 5, pp. 1184–98.

Bijmolt, T. H. A. and Pieters, R. G. M. (2001). Meta-analysis in marketing when studies contain multiple measurements. *Marketing Letters*, **12**, 2, pp. 157–69.

Bleaney, M. and Nishiyama, A. (2004). Income inequality and growth. Does the relationship vary with the income level? *Economics Letters*, **84**, 3, pp. 349–55. ■

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, **87**, 1, pp. 115–43.

Brons, M., Nijkamp, P., Pels, E. and Rietveld, P. (2005). Efficiency of urban public transit: a meta-analysis. *Transportation*, **32**, 1, pp. 1–21.

Brouwer, R. I. H., Langford, I. J., Bateman, I. J. and Turner, R. K. (1999). A meta-analysis of wetland contingent valuation studies. *Regional Environmental Change*, **1**, 1, pp. 47–57.

Button, K. J. (1995). What can meta-analysis tell us about the implications of transport? *Regional Studies*, **29**, pp. 507–17.

Button, K. J. and Kerr, J. (1996). The effectiveness of traffic restraint policies. A simple meta regression model. *International Journal of Transport Economics*, **23**, pp. 213–25.

Button, K. J. and Rietveld, P. (2000). A meta-analysis of the impact of infrastructure policy on regional development. In H. Kohno, P. Nijkamp and J. Poot (eds.), *Regional Cohesion and Competition in the Age of Globalization*. Cheltenham: Edward Elgar.

[23] Studies included in the meta-analysis are marked with the symbol ■.

Campano, F. and Salvatore, D. (2006). *Income Distribution*. Oxford: Oxford University Press.
Card, D. and Krueger, A. B. (1995). Time-series minimum wage studies: a meta-analysis. *American Economic Review*, **85**, pp. 238–43.
Castelló, A. (2004). A reassessment of the relationship between inequality and growth: what human capital inequality data say? IVIE Working Paper no. 15. ■
Castelló, A. and Domenech, R. (2002). Human capital inequality and economic growth: some new evidence. *Economic Journal*, **112**, pp. 187–200. ■
Clarke, G. R. G. (1995). More evidence on income distribution and growth. *Journal of Development Economics*, **47**, pp. 403–27. ■
Cowell, F. A. (1995). *Measuring Inequality* (2nd edn.). London: Prentice Hall.
de la Croix, D. and Doepke, M. (2003). Inequality and growth: why differential fertility matters. *American Economic Review*, **3**, 4, pp. 1091–113. ■
Deininger, K. and Olinto, P. (1998). Asset distribution, inequality, and growth. World Bank Policy Research Working Paper no. 2375. ■
Deininger, K. and Squire, L. (1996). A new data set measuring income inequality. *World Bank Economic Review*, **10**, 3, pp. 563–91.
Deininger, K. and Squire, L. (1998). New ways of looking at old issues: inequality and growth. *Journal of Development Economics*, **57**, pp. 259–87. ■
Dobson, S., Ramlogan, C. and Strobl, E. (2006). Why do rates of β-convergence differ? A meta-regression analysis. *Scottish Journal of Political Economy*, **53**, 2, pp. 153–73.
Egger, M., Smith, G. D., Scheider, M. and Minder, C. (1997). Bias in meta-analysis detected by simple graphical test. *British Medical Journal*, **316**, pp. 629–34.
Figini, P. (1999). Inequality and growth revisited. Trinity College Paper Series no. 99/2. ■
Forbes, K. J. (2000). A reassessment of the relationship between inequality and growth. *American Economic Review*, **90**, 4, pp. 869–87. ■
Galor, O. and Moav, O. (2004). From physical to human capital accumulation: inequality and the process of development. *Review of Economic Studies*, **71**, pp. 1001–26.
Galor, O. and Zang, H. (1997). Fertility, income distribution, and economic growth: theory and cross-country evidence. *Japan and the World Economy*, **9**, 2, pp. 197–229. ■
Galor, O. and Zeira, J. (1993). Income distribution and macroeconomics. *Review of Economic Studies*, **60**, 1, pp. 35–52.
Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, **5**, 10), pp. 3–8.
Goldstein, H. (1995). *Multilevel Statistical Models* (2nd edn.). London: Edward Arnold.
Goldstein, H., Yang, M., Omar, R., Turner, R. and Thompson, S. (2000). Meta-analysis using multilevel models with an application to the study of class size effects. *Applied Statistics*, **49**, 3, pp. 399–412.
Gorg, H. and Strobl, E. (2001). Multinational companies and productivity spillovers: a meta-analysis. *Economic Journal*, **111**, pp. 723–39.
Greene, W. H. (2000). *Econometric Analysis* (4th edn.). Upper Saddle River, NJ: Prentice Hall International.
Gylfason, T. and Zoega, G. (2003a). Inequality and economic growth: do natural resources matter? In T. Eicher and S. Turnovsky (eds.), *Growth and Inequality: Theory and Policy Implications*. Cambridge, MA: MIT Press. ■
Gylfason, T. and Zoega, G. (2003b). Education, social equality and economic growth: a view of the landscape. *CESifo Economic Studies*, **49**, 4, pp. 557–79. ■
Hedges, L. V. and Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. New York: Academic Press.
Iradian, G. (2005). Inequality, poverty, and growth: cross-country evidence. International Monetary Fund (IMF) Working Paper no. WP/05/28. ■
Keefer, P. and Knack, S. (2002). Polarization, politics and property rights: links between inequality and growth. *Public Choice*, **111**, pp. 127–54. ■
Kenworthy, L. (2004). An equality-growth trade-off? In L. Kenworthy ed., *Egalitarian Capitalism. Jobs, Incomes, and Growth in Affluent Countries*. New York: Russell Sage Foundation. ■
Khoo, L. and Dennis, B. (1999). Income inequality, fertility choice, and economic growth: theory and evidence. Harvard University Development Discussion Papers no. 687. ■
Knell, M. (1999). Social comparisons, inequality, and growth. *Journal of Institutional and Theoretical Economics*, **155**, 4, pp. 664–95. ■

KNOWLES, S. (2005). Inequality and economic growth: the empirical relationship reconsidered in the light of comparable data. *Journal of Development Studies*, **41**, 1, pp. 135–59. ■
KUZNETS, S. (1955). Economic growth and income inequality. *American Economic Review*, **45**, pp. 1–28.
LARRAIN, F. B. and VERGARA, R. M. (1997). Income distribution, investment, and growth. Harvard University Development Discussion Papers no. 596. ■
LEONI, T. and POLLAN, W. (2003). The impact of inequality on economic growth. WIFO Working Paper no. 211.
LI, H. and ZOU, H. (1998). Income inequality is not harmful for growth: theory and evidence. *Review of Development Economics*, **2**, 3, pp. 318–34. ■
LITSCHIG, S. (2005). Inequality, institutions and growth: empirical evidence from participatory budgeting experiences in Brazil. Paper presented at the X Spring Meeting of Young Economists, Geneve, April 2005. ■
LONGHI, S., NIJKAMP, P. and POOT, J. (2005). A meta-analytic assessment of the effect of immigration on wages. *Journal of Economic Surveys*, **19**, 3, pp. 451–77.
MBABAZI, J., MORRISSEY, O. and MILNER, C. (2001). Are inequality and trade liberalization influences on growth and poverty? WIDER Discussion Paper no. 2001/132. ■
MIRRLEES, J. A. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies*, **38**, pp. 175–208.
NAHUIS, R. and DE GROOT, H. L. F. (2003). Rising skill premia: you ain't seen nothing yet? CPB Discussion Paper no. 20.
ODEDOKUN, M. O. and ROUND, J. I. (2001). Determinants of income inequality and its effects on economic growth. WIDER Discussion Paper no. 2001/103. ■
PANIZZA, U. (2002). Income inequality and economic growth: evidence from American data. *Journal of Economic Growth*, **7**, pp. 25–41. ■
PARTRIDGE, M. D. (2005). Does income distribution affect U.S. state economic growth? *Journal of Regional Science*, **45**, 2, pp. 363–94. ■
PEROTTI, R. (1996). Growth, income distribution, and democracy: what data say. *Journal of Economic Growth*, **1**, pp. 149–87.
PERSSON, T. and TABELLINI, G. (1991). Is inequality harmful for growth? Theory and evidence. CEPR Working Paper no. 581. ■
PERSSON, T. and TABELLINI, G. (1994). Is inequality harmful for growth? *American Economic Review*, **84**, 3, pp. 600–21.
RAUDENBUSH, S. W. and BRYK, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd edn.). Newbury Park, CA: Sage Publications.
REHME, G. (2002a). (Re-)Distribution of personal incomes, education and economic performance across countries. Luxembourg Income Study Working Paper no. 299. ■
REHME, G. (2002b). Education, economic growth and personal income inequality across (rich) countries. Luxembourg Income Study Working Paper no. 300. ■
REHME, G. (2006). Education, economic growth and measured income inequality. Luxembourg Income Study Working Paper no. 428.
ROSE, A. K. (2005). A meta-analysis of the effect of common currencies on international trade. *Journal of Economic Surveys*, **19**, 3, pp. 347–65.
ROSENTHAL, R. (1991). *Meta-Analytic Procedures for Social Research*. (revised eds.). Newbury Park, CA: Sage Publications.
RUTTER, C. M. and GATSONIS, C. A. (2001). A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Statistics in Medicine*, **20**, pp. 2865–84.
SALVATORE, D. (2007). Income distribution, special issue. *Journal of Policy Modeling*, **29**, 4, pp. 541–655.
SCHIPPER, Y. and HOOGEVEEN, J. G. (2005). Which inequality matters? Growth evidence based on small area welfare estimates in Uganda. World Bank Policy Research Working Paper no. 3592. ■
STANLEY, T. D. (1998). New wine in old bottles: a meta-analysis of Ricardian equivalence. *Southern Economic Journal*, **64**, pp. 713–27.
STANLEY, T. D. (2001). Wheat from chaff: meta-analysis as quantitative literature review. *Journal of Economic Perspectives*, **15**, pp. 131–50.
STANLEY, T. D. (2004). Does unemployment hysteresis falsify the natural rate hypothesis? A meta-regression analysis. *Journal of Economic Surveys*, **18**, 4, pp. 589–612.
STANLEY, T. D. (2005). Beyond publication bias. *Journal of Economic Surveys*, **19**, 3, pp. 309–37.
SUTTON, A. J., ABRAMS, K. R., JONES, D. R., SHELDON, T. A. and SONG, F. (2000). *Methods for Meta-Analysis in Medical Research*. New York: John Wiley and Sons.

SZEKELY, M. and HILGERT, M. (1999). What's behind the inequality we measure: an investigation using Latin American data. Inter-American Development Bank Working Paper no. 409. ■
TANNINEN, H. (1999). Income inequality, government expenditures and growth. *Applied Economics*, **31**, pp. 1109–17. ■
TEMPLE, J. (1999). The new growth evidence. *Journal of Economic Literature*, **37**, 1, pp. 112–56.
VOITCHOVSKY, S. (2005). Does the profile of income inequality matter for economic growth? Distinguishing between the effects of inequality in different Parts of the income distribution. *Journal of Economic Growth*, **10**, 3, pp. 273–96. ■
WARDMAN, M. (2001). A review of British evidence on time and service quality valuations. *Transportation Research, Part E*, **37**, pp. 107–28.
ZHU, D. (2001). Inequality, credit market imperfection, segmentation and economic growth. CentER Discussion Paper no. 2001–58. ■

Date of receipt of final manuscript: 20 March 2008